View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Institutional Repository of the Freie Universität Berlin

FREIE UNIVERSITÄT BERLIN

NONSMOOTH SCHUR–NEWTON METHODS FOR VECTOR-VALUED CAHN–HILLIARD EQUATIONS

CARSTEN GRÄSER, RALF KORNHUBER, AND ULI SACK

Preprint A /01/2013

FACHBEREICH MATHEMATIK UND INFORMATIK
SERIE A • MATHEMATIK

NONSMOOTH SCHUR–NEWTON METHODS FOR VECTOR-VALUED CAHN–HILLIARD EQUATIONS

CARSTEN GRÄSER, RALF KORNHUBER, AND ULI SACK

Abstract. We present globally convergent nonsmooth Schur–Newton methods for the solution of discrete vector-valued Cahn–Hilliard equations with logarithmic and obstacle potentials. The method solves the nonlinear set-valued saddle-point problems as arising from discretization by implicit Euler methods in time and first order finite elements in space without regularization. Efficiency and robustness of the convergence speed for vanishing temperature is illustrated by numerical experiments.

1. Introduction

We consider a vector-valued Cahn–Hilliard system with logarithmic free energy for positive temperature θ and the associated deep quench limit model for $\theta = 0$ describing the isothermal decomposition and coarsening of multicomponent alloys [15, 16, 35]. Existence and uniqueness is shown in the pioneering paper of Elliott and Luckhaus [15]. Discretization by an implicit Euler method in time and by piecewise affine finite elements in space is suggested and analyzed by Blowey et al. [6] and Barrett and Blowey [1] for the logarithmic potential and by Barrett and Blowey [2] in the deep quench limit. Besides the fully implicit Euler method, a globally stable semi-implicit variant taking the concave terms of the logarithmic or obstacle potential explicitly is also considered.

While the numerical analysis of vector-valued Cahn–Hilliard equations with logarithmic free energy and the associated deep quench limit model is well developed, the fast and robust numerical solution of large-scale algebraic systems arising in each time step still seems to be open. Blowey et al. [6] use a nonlinear Gauß–Seidel-type iteration and Barrett and Blowey [1, 2] adapt a splitting method of Lions and Mercier [34] which can be regarded as a type of alternating direction scheme. Both approaches suffer from severe mesh dependence. Boyanova and Neytcheva [8] suggest an inexact Newton method as applicable to polynomial or regularized logarithmic free energies, but do not address the influence of the regularization parameter on the convergence speed. Kim and Kang [31] apply a full approximation storage (FAS) multigrid algorithm to a ternary system with polynomial free energy but neither provide a theoretical convergence result nor discuss the convergence speed of their method.

In this paper we present so-called *nonsmooth Schur–Newton methods* for the spatial problems arising from the discretizations suggested in [1, 2, 6]. Though our approach is applicable to both implicit and semi-implicit time discretizations (see [19,

Key words and phrases. phase field models, variational inequalities, finite elements, convex minimization, descent methods, multigrid methods.

This work was supported by the DFG Research Center Matheon.

Section 3.4.2]), we concentrate on the semi-implicit variant for ease of presentation. Nonsmooth Schur–Newton methods have been introduced, analyzed, and assessed numerically for discretized binary Cahn–Hilliard equations with obstacle potential [22] or logarithmic potential [19, 20]. They can be regarded as gradient-related descent methods for the Schur complement formulation of set-valued saddle point problems, as a preconditioned Uzawa iteration, or as generalizations of well-known primal–dual active set methods [18, 28]. No kind of regularization is involved. The presented extension to the vector-valued case is robust in the sense that global convergence holds for all temperatures $\theta \geq 0$. Moreover, numerical experiments illustrate that the convergence speed is hardly affected by temperature or even by the number of components. Our numerical computations also indicate mesh independent convergence for initial iterates as obtained by nested iteration [27, Chapter 5]. Theoretical validation is the subject of current research.

The paper is organized as follows. After presenting the continuous problem and its discretization, Section 3 concentrates on a unified formulation of the spatial problems that includes both the logarithmic potential and the deep quench limit in terms of a variational inequality. We show existence, uniqueness, and equivalence to a, generally set-valued, nonlinear saddle point problem. We derive nonsmooth Schur–Newton methods for this problem class, analyze their convergence and discuss some algorithmic issues in Section 4. Numerical experiments, as reported in Section 5, illustrate efficiency and robustness of our approach.

2. Vector-valued Cahn–Hilliard equations

We consider phase separation in isothermal multi-component systems on a polygonal (polyhedral) domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$. The concentrations of the different constituents $i = 1, \dots, N$ at $(x, t) \in \Omega \times [0, T_0]$, $T_0 > 0$, are represented by the components $u_i(x, t)$ of the order parameter $u = (u_1, \dots, u_N)^T$.

Throughout the following, we will make use of the Euclidean scalar product $v \cdot w$ with associated norm $|\cdot|$ in Euclidean vector spaces, of the canonical scalar product $(\cdot, \cdot)$ in $L^2(\Omega)$, of the scalar product

$$(v, w) = \int_\Omega v \cdot w \, dx$$

in $L^2(\Omega)^N$ with canonical norm $\|\cdot\|_0$, and of the scalar product

$$(v, w)_1 = (v, w) + (\nabla v, \nabla w), \qquad (\nabla v, \nabla w) = \sum_{i=1}^N \int_\Omega \nabla v_i \cdot \nabla w_i \, dx$$

in $H^1(\Omega)^N$ with canonical norm $\|\cdot\|_1$ and semi norm $|\cdot|_1^2 = (\nabla\cdot, \nabla\cdot)$. Generic constants are denoted by c, C and can have different values at different occurrences.

The order parameter u satisfies the constraints

$$u(x, t) \in G = \{v \in \mathbb{R}^N \mid v_i \geq 0, \ \textstyle\sum_{i=1}^N v_i = 1\} \qquad \forall (x, t) \in \Omega \times [0, T_0],$$

because concentrations are non-negative and add up to unity. The closed convex set $G \subset \mathbb{R}^N$ is often called *Gibbs simplex*. Note that the indicator function χ_G, defined by $\chi_G(u) = 0$ for $u \in G$ and $\chi_G(u) = +\infty$ for $u \notin G$ can be decomposed

according to

$$\chi_G(u) = \sum_{i=1}^{N} \chi_{[0,\infty)}(u_i) + \chi_{V_1}(u), \qquad V_1 = \{v \in \mathbb{R}^N \mid \textstyle\sum_{i=1}^N v_i = 1\},$$

with $\chi_{[0,\infty)}$ and χ_{V_1} denoting the indicator functions of $[0,\infty)$ and V_1, respectively. We assume that the Ginzburg–Landau total free energy of our system takes the form

$$\mathcal{E}(u) = \int_\Omega \frac{\varepsilon}{2} \sum_{i=1}^{N} |\nabla u_i|^2 + \frac{1}{\varepsilon}\Psi(u)\, dx \tag{2.1}$$

with fixed interface parameter $\varepsilon > 0$. While the quadratic interfacial energy is penalizing steep gradients, the free energy Ψ gives rise to phase separation. We concentrate on a multi-phase version of the well-known logarithmic free energy [1, 6]. More precisely, $\Psi = \Psi_\theta$ is given by

$$\Psi_\theta(u) = \Phi_\theta(u) + \chi_{V_1}(u) + \tfrac{1}{2} Ku \cdot u \tag{2.2}$$

with the convex function

$$\Phi_\theta(u) = \begin{cases} \displaystyle\sum_{i=1}^{N} \theta u_i \ln(u_i), & \text{for } \theta > 0, \\ \displaystyle\sum_{i=1}^{N} \chi_{[0,\infty)}(u_i), & \text{for } \theta = 0, \end{cases} \tag{2.3}$$

and a symmetric interaction matrix $K = (K_{ij})_{i,j=1}^N$ (cf. De Fontaine [16]) depending on θ_c. Here, θ and θ_c are denoting absolute and critical temperature, respectively.

For $\theta < \theta_c$, we assume that Ψ_θ has exactly N distinct local minima on G, corresponding to almost pure components $i = 1, \dots, N$. For example, this is achieved by choosing the interaction matrix

$$K = \theta_c N(1 - \delta_{ij})_{i,j=1}^N \qquad \text{(Kronecker-}\delta\text{)} \tag{2.4}$$

which means that the interaction of all different components is equal and no self-interaction occurs. In the deep quench limit $\theta = 0$, we then obtain the classical obstacle potential (cf. Blowey & Barrett [2])

$$\Psi_0(u) = \chi_G(u) + \theta_c \tfrac{N}{2} \sum_{i=1}^{N} u_i(1 - u_i).$$

For $\theta > 0$ and $N = 2$ the well-known logarithmic free energy

$$\Psi_\theta(\tilde{u}) = \tfrac{1}{2}\theta[(1+\tilde{u})\ln(\tfrac{1+\tilde{u}}{2}) + (1-\tilde{u})\ln(\tfrac{1-\tilde{u}}{2})] + \tfrac{1}{2}\theta_c(1-\tilde{u}^2)$$

of the scalar order parameter $\tilde{u} := u_2 - u_1 \in [-1, 1]$ is recovered in this way. In the shallow quench, i.e., for $\theta \approx \theta_c$, polynomial free energies generalizing the quartic potential $(1-\tilde{u}^2)^2$ provide good approximations of Ψ_θ (cf. Steinbach et al. [39]). As polynomials are defined everywhere, the contributions from the non-differentiable indicator function $\chi_{[0,\infty)}$ are usually skipped in this case.

For $\theta > 0$ the vector-valued Cahn–Hilliard equation

$$\begin{aligned} u_t &= L\Delta w, \\ w &= -\varepsilon^2 \Delta u + P\Psi_\theta'(u) \end{aligned} \tag{2.5}$$

is obtained by postulating that u satisfies a conservation law with flux $-L\nabla w$ and w given by the derivative of the total free energy $\mathcal{E}$ defined in (2.1). In this way, the mass of all components u_i is conserved and (2.5) is thermodynamically consistent in the sense that $\mathcal{E}$ decreases monotonically throughout the evolution. We assume that the matrix $L \in \mathbb{R}^{N\times N}$ is symmetric and positive semi-definite with one-dimensional kernel spanned by $\mathbf{1} = (1,\dots,1)^\top \in \mathbb{R}^N$. The latter condition accounts for the fact that $u \cdot \mathbf{1} \equiv 1$ and therefore $u_t \cdot \mathbf{1} \equiv 0$. The orthogonal projection

$$P = I - \tfrac{1}{N}(\mathbf{1},\dots,\mathbf{1}) \in \mathbb{R}^{N\times N},$$

maps $\mathbb{R}^N$ onto the linear subspace

$$V_0 = \{v \in \mathbb{R}^N \mid \textstyle\sum_{i=1}^N v_i = 0\} \subset \mathbb{R}^N.$$

It accounts for the fact that admissible variations of $u(x,t) \in V_1$ must be in V_0.

In the singular deep quench limit $\theta = 0$, the second equation in (2.5) becomes

$$w \in -\varepsilon^2 \Delta u + P\partial\Psi_0(u). \tag{2.6}$$

We assume that the initial condition $u_0 \in H^1(\Omega)$ satisfies $u_0(x) \in V_1$ for almost all $x \in \Omega$ and the componentwise inequality

$$0 < \frac{1}{|\Omega|}\int_\Omega u_0(x)\,dx < \mathbf{1} \tag{2.7}$$

which means that the given problem is non-degenerate in the sense that exactly N different components are actually occurring. We further impose Neumann boundary conditions for u and w so that mass conservation

$$\int_\Omega u(x,t)\,dx = \int_\Omega u_0(x)\,dx,$$

and

$$(I-P)u(x,t) = \tfrac{1}{N}\mathbf{1}, \qquad (I-P)w(x,t) = 0$$

holds for almost all $x \in \Omega$ and $t > 0$. On these assumptions, existence and uniqueness was shown by Elliott and Luckhaus [15] for $\theta \geq 0$. For sharp interface limits, we refer to Bronsard et al. [9].

3. Discretization

3.1. **Semi-implicit time discretization and finite elements.** Let us first consider the case $\theta > 0$. Then time discretization of a weak formulation of (2.5) by the implicit Euler scheme and subsequent finite element discretization leads to spatial problems of the form

(FI) Find $u \in \mathcal{G}, w \in \mathcal{S}^N$ such that

$$\begin{aligned} \varepsilon^2\,(\nabla u, \nabla v) + (P\Phi_\theta'(u), v)^{\mathcal{T}} + (PKu, v) &= (w, v) && \forall v \in \mathcal{S}^N, \\ (u, v) + \tau\,(L\nabla w, \nabla v) &= \left(u^{\mathrm{old}}, v\right) && \forall v \in \mathcal{S}^N. \end{aligned}$$

Here, $\tau > 0$ denotes the uniform time step size, u and w stand for the finite element approximations of order parameter and chemical potential in the given time step, respectively, and u^{old} denotes the approximate order parameter in the preceding time step. The finite element space $\mathcal{S}^N$ is the tensor product of scalar, piecewise affine finite elements

$$\mathcal{S} = \{v \in C(\overline{\Omega}) \mid v|_T \text{ is affine } \forall T \in \mathcal{T}\}$$

as induced by a simplicial partition $\mathcal{T}$ of Ω. We assume that $\mathcal{T} = \mathcal{T}_j$ together with an underlying hierarchy $\mathcal{T}_0, \dots, \mathcal{T}_j$ is obtained by successive adaptive refinement of a conforming, intentionally coarse partition $\mathcal{T}_0$. During this refinement process, so-called "hanging" vertices are allowed to occur at the midpoints of certain edges. Each function $v \in \mathcal{S}$ is characterized by its values in $p \in \mathcal{N}$, the set of all "non-hanging" vertices of simplices $T \in \mathcal{T}$. Therefore, $\mathcal{S}$ is spanned by the nodal basis $\lambda_p \in \mathcal{S}$, $p \in \mathcal{N}$, defined by the condition $\lambda_p(q) = \delta_{p,q}\ \forall q \in \mathcal{N}$. We refer to [19] for details.

We have also used the closed convex subset

$$\mathcal{G} = \{v \in \mathcal{S}^N \mid v(p) \in G \ \ \forall p \in \mathcal{N}\},$$

and the lumped L^2 scalar product

$$(u, v)^{\mathcal{T}} = \int_\Omega I^{\mathcal{T}}(u \cdot v)\, dx$$

induced by canonical nodal interpolation $I^{\mathcal{T}} v = \sum_{p\in\mathcal{N}} v(p)\lambda_p$. Note that lumping has been applied only to the nonlinear term $(\Phi'_\theta(u), v)$ in order to separate the unknowns associated with different nodes with respect to nonlinearity. Full lumping, i.e., lumping of all other zero-order terms, is quite common in the literature (cf., e.g. [1, 2, 3, 4, 6]) but is avoided here, because it would either destroy symmetry or mass conservation if the underlying grids have changed from the preceding time step to the given time step [19, Section 3.4.3], [25, Section 5].

All reasoning to be presented below extends a fully lumped version of (FI) as proposed and analyzed by Blowey et al. [6] and Barrett and Blowey [1]. Existence and uniqueness of corresponding discrete solutions has been shown in [6, Theorem 2.4] for the fully lumped version on the time step constraint

$$\tau < 4\varepsilon^2/(\lambda_K^2 \|L\|).$$

Here, λ_K denotes the largest positive eigenvalue of K and $\|L\|$ stands for the spectral norm of L. For example, for K taken from (2.4) the time step τ has to satisfy $\tau < 4\varepsilon^2/(\theta_c^2 N^2 (N-1)^2 \|L\|)$. In order to avoid such severe stability restrictions, the expanding linear part K of $\Psi'_\theta = \Phi'_\theta + K$ on $\mathcal{G}$ is often discretized explicitly (cf., e.g., [6, 7, 23, 30]). More precisely, Ku is replaced by $Ku^{\mathrm{old}} + (P-I)Ku$ providing the unconditionally stable semi-implicit scheme

(SI) Find $u \in \mathcal{G}, w \in \mathcal{S}^N$ such that

$$\begin{aligned} \varepsilon^2 (\nabla u, \nabla v) + (P\Phi'_\theta(u), v)^{\mathcal{T}} + \big(Ku^{\mathrm{old}} + (P-I)Ku, v\big) &= (w, v) \qquad \forall v \in \mathcal{S}^N, \\ (u, v) + \tau\, (L\nabla w, \nabla v) &= \big(u^{\mathrm{old}}, v\big) \quad \forall v \in \mathcal{S}^N. \end{aligned}$$

By construction $u(p) \in G\ \forall p \in \mathcal{N}$. Orthogonality of P with respect to the Euclidean scalar product immediately provides $(I-P)w \equiv 0$ for the fully implicit discretization (FI). For the semi-implicit version (SI), testing with $v = \lambda_p \mathbf{1}$ and a short calculation shows

$$(w \cdot \mathbf{1}, \lambda_p) = (K(u^{\mathrm{old}} - u) \cdot \mathbf{1}, \lambda_p). \tag{3.1}$$

Hence $(I-P)w \equiv 0$ is no longer true in general but only holds in special cases, e.g. for the choice (2.4) of K.

3.2. **Unified formulation of spatial problems for $\theta \geq 0$.** The occurrence of P in the projected derivative $P\Phi'_\theta(\cdot)$ in the discretizations (FI) and (SI) prevents a direct reformulation as a variational inequality that would allow to pass to the deep quench limit $\theta = 0$. Utilizing $(I-P)w \equiv 0$, such a formulation can be easily obtained for the fully implicit version (FI). We therefore concentrate on the semi-implicit variant and first introduce the (affine) subspaces

$$\mathcal{S}^N_r = \{v \in \mathcal{S}^N \mid v(p) \in V_r \ \forall p \in \mathcal{N}\}, \qquad r \in \{0,1\}.$$

Using the reduced test space $\mathcal{S}^N_0 \subset \mathcal{S}^N$ in the first equation of (SI), we obtain

$$\varepsilon^2 (\nabla u, \nabla v) + (\Phi'_\theta(u), v)^{\mathcal{T}} + \big(Ku^{\mathrm{old}}, v\big) = (w_0, v) \qquad \forall v \in \mathcal{S}^N_0$$

with the new variable $w_0 = Pw \in \mathcal{S}^N_0$. We now rewrite this equation as a variational inequality and use the reduced test space $\mathcal{S}^N_0 \subset \mathcal{S}^N$ in the second equation of (SI) as well, to obtain

($\widehat{\mathrm{VI}}$) Find $u \in \mathcal{S}^N_1, w \in \mathcal{S}^N_0$ such that

$$\begin{aligned} \varepsilon^2 (\nabla u, \nabla(v-u)) + \phi^{\mathcal{T}}_\theta(v) - \phi^{\mathcal{T}}_\theta(u) & \\ - (w_0, v-u) \geq -\big(Ku^{\mathrm{old}}, v-u\big) & \qquad \forall v \in \mathcal{S}^N_1, \\ (u,v) + \tau (L\nabla w_0, \nabla v) = \big(u^{\mathrm{old}}, v\big) & \qquad \forall v \in \mathcal{S}^N_0. \end{aligned}$$

Note that the lumped nonlinearity $(\Phi'_\theta(u), v)^{\mathcal{T}}$ gives rise to the nonlinear functional $\phi^{\mathcal{T}}_\theta$, defined by

$$\phi^{\mathcal{T}}_\theta(v) = \begin{cases} \displaystyle\int_\Omega I^{\mathcal{T}}\big(\Phi_\theta(v)\big)\, dx, & \text{if } v \geq 0 \\ +\infty, & \text{else} \end{cases}$$

The variational problem ($\widehat{\mathrm{VI}}$) has the advantage that it allows for a straightforward extension to the deep quench limit $\theta = 0$. In this case, $\phi^{\mathcal{T}}_0$ just accounts for the inequality constraints $u_i \geq 0$. For positive temperature, the variational formulation ($\widehat{\mathrm{VI}}$) is equivalent to (SI) in the following sense.

Proposition 3.1. *Let $\theta > 0$. If (u,w) is a solution of (SI) then (u, Pw) is a solution of ($\widehat{\mathrm{VI}}$), and if (u, w_0) is a solution of ($\widehat{\mathrm{VI}}$), then there is a solution (u,w) of (SI) with $w_0 = Pw$.*

Proof. Let (u,w) be a solution of (SI). Then (u, Pw) is a solution of ($\widehat{\mathrm{VI}}$) by construction.

Now let (u, w_0) be a solution of ($\widehat{\mathrm{VI}}$). Then we use the decomposition

$$v = Pv + (I-P)v = v_0 + v_1 \mathbf{1}, \quad v_0 = Pv \in \mathcal{S}^N_0, \quad v_1 = \tfrac{1}{N} v \cdot \mathbf{1}$$

of all $v \in \mathcal{S}^N$ to define (cf. (3.1))

$$w = w_0 + w_1 \mathbf{1}, \qquad (w_1, \lambda_p) = \tfrac{1}{N}(K(u^{\mathrm{old}} - u) \cdot \mathbf{1}, \lambda_p) \quad \forall p \in \mathcal{N}.$$

Note that w is well defined because the mass matrix $\big((\lambda_p, \lambda_q)\big)_{p,q \in \mathcal{N}}$ is invertible. Now, exploiting the orthogonality of P and the properties of L, it is easily checked that u, w solve (SI). □

3.3. **Weak formulation of affine constraints.** In order to simplify the algebraic solution, we now derive a version of (SI) that incorporates the affine constraints $u(p) \cdot \mathbf{1} = 1$ in the weak form

$$(u, \mathbf{1}v) = (u \cdot \mathbf{1}, v) = (1, v) \qquad \forall v \in \mathcal{S} \tag{3.2}$$

and not in the strong form $u \in \mathcal{S}_1^N$. Introducing the Lagrange multiplier $\mathbf{1}\eta \in \mathcal{S}^N$ associated with (3.2), the solution of ($\widehat{\text{VI}}$) amounts to find $u \in \mathcal{S}^N, w_0 \in \mathcal{S}_0^N, \eta \in \mathcal{S}$ such that

$$\begin{aligned}
\varepsilon^2 (\nabla u, \nabla(v-u)) + \phi_\theta^{\mathcal{T}}(v) - \phi_\theta^{\mathcal{T}}(u) & \\
- (w_0 + \mathbf{1}\eta, v-u) \geq - \left(K u^{\text{old}}, v-u\right) & \qquad \forall v \in \mathcal{S}^N, \\
- (u, v) - \tau (L \nabla w_0, \nabla v) = - \left(u^{\text{old}}, v\right) & \qquad \forall v \in \mathcal{S}_0^N \\
- (u, \mathbf{1}v) = - \left(u^{\text{old}}, \mathbf{1}v\right) & \qquad \forall v \in \mathcal{S}.
\end{aligned}$$

For the deep quench limit $\theta = 0$ a fully implicit and fully lumped version of this discretization has been suggested and analyzed by Barrett and Blowey [2].

Introducing the new variable $\tilde{w} = w_0 + \mathbf{1}\eta$, adding the last two equations, and using $L\mathbf{1} = 0$, we see that this problem is equivalent to finding $u \in \mathcal{S}^N, \tilde{w} \in \mathcal{S}^N$ such that

$$\begin{aligned}
\varepsilon^2 (\nabla u, \nabla(v-u)) + \phi_\theta^{\mathcal{T}}(v) - \phi_\theta^{\mathcal{T}}(u) & \\
- (\tilde{w}, v-u) \geq - \left(K u^{\text{old}}, v-u\right) & \qquad \forall v \in \mathcal{S}^N, \\
- (u, v) - \tau (L \nabla \tilde{w}, \nabla v) = - \left(u^{\text{old}}, v\right) & \qquad \forall v \in \mathcal{S}^N.
\end{aligned}$$

In the final step, we enforce coercivity of the primal operator by exploiting mass conservation

$$\int_\Omega u_i(x)\, dx = \int_\Omega u_i^{\text{old}}(x)\, dx.$$

in a similar way as introduced in [17, 19, 22]. More precisely, we add the equality

$$\varepsilon^2 (u, \mathbf{1})(v-u, \mathbf{1}) = \varepsilon^2 (u^{\text{old}}, \mathbf{1})(v-u, \mathbf{1}) \qquad \forall v \in \mathcal{S}^N.$$

to the variational inequality above, to obtain the final form of the spatial problem

(VI) Find $(u, w) \in \mathcal{S}^N \times \mathcal{S}^N$ such that

$$\begin{aligned}
\varepsilon^2 (\nabla u, \nabla(v-u)) + \varepsilon^2 \int_\Omega u\, dx \cdot \int_\Omega v - u\, dx + \phi_\theta^{\mathcal{T}}(v) - \phi_\theta^{\mathcal{T}}(u) & \\
- (w, v-u) \geq \varepsilon^2 \int_\Omega u^{\text{old}}\, dx \cdot \int_\Omega v - u\, dx - \left(K u^{\text{old}}, v-u\right) & \quad \forall v \in \mathcal{S}^N, \\
- (u, v) - \tau (L \nabla w, \nabla v) = - \left(u^{\text{old}}, v\right) & \quad \forall v \in \mathcal{S}^N.
\end{aligned}$$

In the light of the above considerations and Proposition 3.1, the formulations (SI) and (VI) are equivalent for positive temperature in the following sense.

Proposition 3.2. *Let $\theta > 0$. If (u, w) is a solution of (SI) then there is a solution $(u, \tilde{w})$ of (VI) satisfying $Pw = P\tilde{w}$ and vice versa.*

The common idea behind the above reformulations is to use the part $(I-P)w \in \mathcal{S}$ as a kind of dustbin, e.g., for the Lagrange parameter in $\eta \in \mathcal{S}$.

3.4. **Existence and uniqueness of discrete solutions.** The variational problem (VI) is equivalent to a saddle point problem to find $(u,w) \in \mathcal{S}^N \times \mathcal{S}^N$ such that

$$\mathcal{L}(u,w) = \inf_{v\in\mathcal{S}^N} \sup_{z\in\mathcal{S}^N} \mathcal{L}(v,z) \tag{3.3}$$

holds for the Lagrangian

$$\mathcal{L}(v,z) = \mathcal{J}(v) + (u^{\mathrm{old}} - v, z) - \frac{\tau}{2}\,(L\nabla z, \nabla z)$$

involving the coercive, convex, lower semi-continuous energy functional

$$\mathcal{J}(v) = \frac{\varepsilon^2}{2}\,(\nabla v, \nabla v) + \frac{\varepsilon^2}{2}\left|\int_\Omega v - u^{\mathrm{old}}\,dx\right|^2 + \phi_\theta^{\mathcal{T}}(v) + \big(Ku^{\mathrm{old}}, v\big)\,.$$

The Lagrangian $\mathcal{L}$ has finite values on the closed, convex set

$$\mathrm{dom}(\mathcal{J}) \times \mathcal{S}^N = \{v \in \mathcal{S}^N \mid v \geq 0\} \times \mathcal{S}^N.$$

In order to show existence of discrete solutions of (3.3), we will make use of a dual problem for w only. The key ingredient for solutions of the dual problem is the following lemma.

Lemma 3.1. *The functional $h = -\inf_{v\in\mathrm{dom}(\mathcal{J})} \mathcal{L}(v,\cdot)$ is coercive on $\mathcal{S}^N$.*

Proof. Let $z \in \mathcal{S}^N$ be arbitrary and define a corresponding $v_0 = v^{(1)} + v^{(2)} \in \mathcal{S}^N$ with $v^{(1)}, v^{(2)} \in \mathcal{S}^N$ given by the nodal values

$$v^{(1)}(p) = \tfrac{1}{2}\big(1 + \mathrm{sgn}(z^{(1)})\big), \qquad v^{(2)}(p) = \rho\big(1 + \mathrm{sgn}(\mathbf{1}\cdot z(p))\big)\mathbf{1}, \qquad p \in \mathcal{N}.$$

Here we have used the componentwise mean value of z,

$$z^{(1)} = |\Omega|^{-1}\int_\Omega z\,dx, \qquad |\Omega| = \int_\Omega dx,$$

and some positive $\rho \in \mathbb{R}$ to be specified later. In the light of

$$-\inf_{v\in\mathrm{dom}(\mathcal{J})} \mathcal{L}(v,z) \geq -\mathcal{L}(v_0,z) = -\mathcal{J}(v_0) - (u^{\mathrm{old}} - v_0, z) + \frac{\tau}{2}\,(L\nabla z, \nabla z) \tag{3.4}$$

we now derive an upper bound for $(u^{\mathrm{old}} - v_0, z)$. To this end, we first decompose z according to

$$z = z^{(0)} + z^{(1)}, \qquad z^{(0)} = z - z^{(1)} = z - |\Omega|^{-1}\int_\Omega z\,dx.$$

Utilizing $u^{\mathrm{old}}(p) \in V_1$ and the definition of $z^{(0)}$ and $v^{(1)}$, we then have

$$(u^{\mathrm{old}}, (I-P)z^{(0)}) = \frac{1}{N}(\mathbf{1}, z^{(0)}) = 0 = \rho(\mathbf{1}, z^{(0)}), \qquad (v^{(1)}, z^{(0)}) = 0.$$

These identities and the properties of P provide

$$\begin{aligned}\big(u^{\mathrm{old}} - v_0, z\big) &= \Big(u^{\mathrm{old}}, Pz^{(0)}\Big) + \Big(u^{\mathrm{old}}, (I-P)z^{(0)}\Big) - \Big(v_0, z^{(0)}\Big) \\ &\qquad + \Big(u^{\mathrm{old}} - v^{(1)}, z^{(1)}\Big) - \Big(v^{(2)}, z^{(1)}\Big) \\ &= \Big(u^{\mathrm{old}}, Pz^{(0)}\Big) - \Big(\rho\mathbf{1}, z^{(0)}\Big) + \Big(u^{\mathrm{old}} - v^{(1)}, z^{(1)}\Big) - \Big(v^{(2)}, z\Big) \\ &= \Big(u^{\mathrm{old}}, Pz^{(0)}\Big) + \Big(u^{\mathrm{old}} - \rho\mathbf{1} - v^{(1)}, z^{(1)}\Big) + \Big(\rho\mathbf{1} - v^{(2)}, z\Big)\,.\end{aligned}$$

Using the Poincaré inequality, the first term can be estimated by

$$\left(u^{\mathrm{old}}, Pz^{(0)}\right) \le \|u^{\mathrm{old}}\|_0 \|Pz^{(0)}\|_0 \le C_0 |Pz^{(0)}|_1 = C_0 |Pz|_1 \tag{3.5}$$

with C_0 independent of z. In order to estimate the second term, we now select

$$\rho = \frac{1}{2|\Omega|} \min_{i=1,\dots,N} \int_\Omega u_i^{\mathrm{old}}\, dx > 0$$

and set $\mu_i = |\Omega|^{-1} \int_\Omega u_i^{\mathrm{old}} - \rho\, dx$. Note that $0 < \mu_i < 1$. Investigating the three cases $v_i^{(1)} \in \{0, \frac{1}{2}\}$, using the equivalence of norms on $\mathbb{R}^N$, and that the orthogonal projection has unit norm, we obtain

$$\begin{aligned}\left(u^{\mathrm{old}} - \rho\mathbf{1} - v, z^{(1)}\right) &= \sum_{i=1}^N (\mu_i - v_i^{(1)}) \int_\Omega z_i\, dx = -\sum_{i=1}^N |\mu_i - v_i^{(1)}| \left| \int_\Omega z_i\, dx \right| \\ &\le -c_0 \sum_{i=1}^N \left| \int_\Omega z_i\, dx \right| \le -c_0 \sqrt{N} \left| \int_\Omega z\, dx \right| \le -c_0 \sqrt{N} \left| \int_\Omega Pz\, dx \right|\end{aligned}$$

Here, c_0 is defined by

$$c_0 = \min_{\substack{i=1,\dots,N \\ z_i^{(1)} \neq 0}} |\mu_i - v_i^{(1)}| > 0$$

if there is at least one i such that $z_i^{(1)} = |\Omega|^{-1} |\int_\Omega z_i\, dx| \neq 0$ and $c_0 = 1$ otherwise. In order to treat the third term $\left(\rho\mathbf{1} - v^{(2)}, z\right)$, we utilize the identities

$$\begin{aligned}(\rho\mathbf{1} - v^{(2)}(p)) \cdot z(p) &= -\rho\, \mathrm{sgn}[\mathbf{1} \cdot z(p)] \mathbf{1} \cdot (I - P) z(p), = -\rho |\mathbf{1} \cdot z(p)| \\ |(I - P) z(p)| &= \frac{1}{\sqrt{N}} |\mathbf{1} \cdot z(p)|\end{aligned}$$

to obtain

$$\left(\rho\mathbf{1} - v^{(2)}, z\right) = -\rho \int_\Omega |\mathbf{1} \cdot z|\, dx. = -\rho\sqrt{N} \int_\Omega |(I - P) z|\, dx.$$

Inserting these three estimates and the identity $(L\nabla z, \nabla z) = |Pz|_1^2$ into (3.4), we obtain

$$-\mathcal{L}(v_0, z) \ge C \left(|Pz|_1^2 - |Pz|_1 + \left| \int_\Omega Pz\, dx \right| + \int_\Omega |(I - P) z|\, dx - 1 \right) \tag{3.6}$$

with a constant C independent of z. In order to show that the right hand side of this inequality tends to infinity, if (a suitable norm of) z tends to infinity, observe that Poincaré's inequality yields

$$\begin{aligned}|Pz|_1^2 - |Pz|_1 + \left| \int_\Omega Pz\, dx \right| &\ge |Pz|_1 + \left| \int_\Omega Pz\, dx \right| - 1 \\ &\ge c\, (|Pz|_1 + \|Pz\|_0 - 1) \ge c(\|Pz\|_1 - 1)\end{aligned}$$

with positive c independent of z. Inserting this estimate into (3.6), we finally get

$$-\mathcal{L}(v_0, z) \ge C \left(\|Pz\|_1 + \int_\Omega |(I - P) z| - 1 \right). \tag{3.7}$$

with a constant C independent of z. This concludes the proof. □

Now we are ready to show existence and uniqueness. Here, the condition

$$0 < \int_\Omega u^{\mathrm{old}}\, dx. \tag{3.8}$$

follows from the non-degeneracy condition (2.7) by componentwise mass conservation of (VI).

Theorem 3.1. *Assume that u^{old} satisfies the non-degeneracy condition* (3.8) *Then (VI) has a solution (u, w) with uniquely determined u and $\nabla P w$.*

Proof. Later, we will show in Proposition 4.1 that (i) the functional h defined in Lemma 3.1 is convex, continuous, and finite and (ii) that minimizing h over $\mathcal{S}^N$ is equivalent to (VI). Hence existence of a minimizer w of h and thus of a solution (u, w) of (VI) follows from convexity and continuity stated in Proposition 4.1 together with coercivity stated in Lemma 3.1, see, e.g., [13, Chapter II, Proposition 1.2].

It remains to show that u and $\nabla P w$ are uniquely determined. To this end, let u^1, w^1 and u^2, w^2 be two solutions of (VI). As the variational inequality in (VI) is satisfied by (u^1, w^1) with $v = u^2$ and by (u^2, w^2) with $v = u^1$, we can add the resulting inequalities to obtain

$$\varepsilon^2 \left(\nabla(u^1 - u^2), \nabla(u^1 - u^2)\right) + \varepsilon^2 \int_\Omega (u^1 - u^2) \cdot \int_\Omega (u^1 - u^2) \le \left(w^1 - w^2, u^1 - u^2\right).$$

Similarly, the variational equality in (VI) holds for (u^1, w^1) with $v = w^1 - w^2$ and for (u^2, w^2) with $v = w^2 - w^1$. Adding the resulting equations, we obtain

$$\left(w^1 - w^2, u^1 - u^2\right) = -\tau \left(L\nabla(w^1 - w^2), \nabla(w^1 - w^2)\right) \le 0$$

which yields the assertion. □

Uniqueness of the chemical potential w is available on additional conditions.

Theorem 3.2. *In addition to the non-degeneracy condition in Theorem 3.1, assume that there a subset*

$$\mathcal{B} = \{\eta_{i,j} \mid \eta_{i,j} = e^i - e^j \text{ with } i \neq j \text{ and } u_i(p), u_j(p) > 0 \text{ for some } p \in \mathcal{N}\} \subset V_0 \tag{3.9}$$

such that $\operatorname{span} \mathcal{B} = V_0$. *Then the solution (u, w) of (VI) is unique.*

Proof. Let (u, w^1) and (u, w^2) be solutions of (VI). In a first step, we show $Pw^1 = Pw^2$. As $\nabla P(w^1 - w^2) = 0$ holds according to Theorem 3.1, we have $P(w^1 - w^2) \equiv v_0$ for some vector $v_0 \in V_0$. Now consider some arbitrary $\eta_{i,j} \in \mathcal{B}$ with associated vertex $p \in \mathcal{N}$ such that $u_i(p), u_j(p) > 0$. Then $v_\pm = u \pm \delta \eta_{i,j} \lambda_p \ge 0$ holds for sufficiently small $\delta > 0$. The variational inequality in (VI) is satisfied for (u, w^1) and the test function $v = v_+$ as well as for (u, w^2) and the test function $v = v_-$. We add these inequalities and divide by δ to obtain

$$\frac{\phi_\theta^{\mathcal{T}}(u + \delta\eta_{i,j}\lambda_p) - \phi_\theta^{\mathcal{T}}(u)}{\delta} - \frac{\phi_\theta^{\mathcal{T}}(u) - \phi_\theta^{\mathcal{T}}(u - \delta\eta_{i,j}\lambda_p)}{\delta} - \left(w^1 - w^2, \eta_{i,j}\lambda_p\right) \ge 0.$$

As $u_i(p), u_j(p) > 0$, the scalar function $\xi \mapsto \phi_\theta^{\mathcal{T}}(u + \xi\eta_{i,j}\lambda_p)$ is differentiable in $\xi = 0$. Hence, we can pass to the limit $\delta = 0$ to get

$$-\left(w^1 - w^2, \eta_{i,j}\lambda_p\right) \ge 0.$$

Exchanging the role of w^1 and w^2, we obtain

$$0 = \left(w^1 - w^2, \eta_{i,j}\lambda_p\right) = \left(P(w^1 - w^2), \eta_{i,j}\lambda_p\right) = v_0 \cdot \eta_{i,j} \int_\Omega \lambda_p \, dx. \tag{3.10}$$

As (3.10) holds for all vectors $\eta_{i,j}$ in the spanning subset $\mathcal{B}$, this implies $v_0 = 0$ and thus $Pw^1 = Pw^2$.

In the next step we show that $w^1 - w^2 = (I - P)(w^1 - w^2) = 0$. To this end, first note that, by definition of P, the identity $(I - P)(w^1 - w^2) = w_0\mathbf{1}$ must hold with some scalar function $w_0 \in \mathcal{S}$. In order to show $w_0 = 0$, we fix some arbitrary node $p \in \mathcal{N}$ and select $1 \le i \le N$ such that $u_i(p) > 0$. This is possible, because $u(p) \cdot \mathbf{1} = 1$ holds for all nodes $p \in \mathcal{N}$. Then $v_\pm = u \pm \delta e^i \lambda_p \in \mathcal{S}^N$ satisfies $v_\pm \ge 0$ for sufficiently small $\delta > 0$. Now, proceeding literally as above, we get

$$0 = \left(w^1 - w^2, e^i\lambda_p\right) = \left(w_0\mathbf{1}, e^i\lambda_p\right) = (w_0, \lambda_p)\,.$$

As p was arbitrary, this implies $w_0 = 0$. □

The assumption on $\mathcal{N}$ in Theorem 3.2 essentially means that the discrete interfacial region is rich enough to contain a suitable set of nodal basis functions. This assumption can be replaced by the more instructive, but much stronger condition that at least one of the components that are present at a certain vertex must also be present at each neigboring one. This property can be always achieved by resolving the diffuse interface sufficiently well.

Lemma 3.2. *Assume that u^{old} satisfies the non-degeneracy condition* (3.8) *and that for any pair (p, q) of neighboring vertices we have*

$$\{1 \le i \le N \mid u_i(p) > 0\} \cap \{1 \le i \le N \mid u_i(q) > 0\} \neq \emptyset. \tag{3.11}$$

Then there is set $\mathcal{B} \subset V_0$ of vertices satisfying the assumption in Theorem 3.2.

Proof. It is sufficient to construct subsets $\mathcal{B}_i$, $i = 1, \dots, N$, of the form (3.9), i.e.,

$$\mathcal{B}_i = \{\eta_{k,j} \mid \eta_{k,j} = e^k - e^j \text{ with } k \neq j \text{ and } u_k(p), u_j(p) > 0 \text{ for some } p \in \mathcal{N}\}$$

such that $e^i - e^1 \in \operatorname{span} \mathcal{B}_i$, because then the vectors $e^i - e^1$, $i = 1, \dots, N$, spanning V_0, are contained in the span of

$$\mathcal{B} := \bigcup_{i=1}^N \mathcal{B}_i.$$

Let us consider some fixed $i \in \{1, \dots, N\}$. By the degeneracy condition (3.8) there are vertices q_1, $q_i \in \mathcal{N}$ such that $u_1(q_1) > 0$, $u_i(q_i) > 0$. Since the grid $\mathcal{T}$ is a connect graph, there is a path $p^1, \dots, p^K$ of neighboring vertices with $p^1 = q_1$ and $p^K = q_i$.

We now assign a component $c_k \in \{1, \dots, N\}$ to each p^k in the following way. We start by setting $c_1 = 1$ so that $u_{c_1}(p^1) > 0$ and assume $u_{c_{k-1}}(p^{k-1}) > 0$ for some $k > 1$. Then we keep $c_k := c_{k-1}$, if $u_{c_{k-1}}(p^k) > 0$, i.e., if the component c_{k-1} is still present at the neighboring vertex p^k. If this is not the case, then we switch to a new component c_k with $u_{c_k}(p^k) > 0$ and $u_{c_k}(p^{k-1}) > 0$, i.e. to a new component which is present in both vertices p^k and p^{k-1}. This is possible due to assumption (3.11). Finally we formally set $c_{K+1} = i$ and define

$$\mathcal{B}_i = \{e^{c_k} - e^{c_{k-1}} \mid c_k \neq c_{k-1}, k = 2, \dots, K+1\}$$

By construction, $u_{c_k}(p^{k-1})$, $u_{c_{k-1}}(p^{k-1}) > 0$ holds for all $k = 2, \dots, K+1$ so that $\mathcal{B}_i$ is of the desired form. Moreover, we have $e^i - e^1 \in \operatorname{span} \mathcal{B}_i$ using the telescope sum

$$e^i - e^1 = \sum_{k=2}^{K+1} e^{c_k} - e^{c_{k-1}} \in \operatorname{span} \mathcal{B}_i.$$

□

Obviously, the assumption in Lemma 3.2 can be weakened by applying the same arguments to certain paths of not necessarily neighboring vertices. However, this essentially amounts to a reformulation of the abstract assumption of Theorem 3.2 again.

3.5. **Algebraic formulation.** Now we rewrite the discrete problem (VI) in an algebraic fashion. This will simplify the derivation and convergence analysis of nonsmooth Schur–Newton methods for the iterative solution of (VI) to be presented in the next section. Starting from an enumeration $\mathcal{N} = \{p_1, \dots, p_m\}$ of the $m = \#\mathcal{N}$ vertices, we enumerate the $n = mN$ nodal basis functions of $\mathcal{S}^N$ according to

$$\{\lambda^1, \lambda^2, \dots, \lambda^n\}, \qquad \lambda^{\pi(i,k)} = e^i \lambda_{p_k},\ i = 1, \dots, N,\ k = 1, \dots m,$$

utilizing the bijective index map $\pi\{1, \dots, N\} \times \{1, \dots, m\} \to \{1, \dots, n\}$ defined by $\pi(i,k) = i + N(k-1)$. Utilizing the canonical isomorphism $\mathcal{S}^N \ni v \mapsto V = (V_i) \in \mathbb{R}^n$ induced by the basis representation

$$v = \sum_{i=1}^{n} V_i \lambda^i, \qquad v \in \mathcal{S}^N,$$

the solution of (IV) amounts to find U, $W \in \mathbb{R}^n$ such that

(3.12)
$$\begin{aligned} AU \cdot (V - U) + \varphi(V) - \varphi(U) + BW \cdot (V - U) \geq \quad & f \cdot (V - U) \quad \forall V \in \mathbb{R}^n \\ BU - CW = \quad & g \end{aligned}$$

Here we have used the matrices $A = (A_{ij})$, $B = (B_{ij})$, $C = (C_{ij}) \in \mathbb{R}^{n,n}$ given by

(3.13)
$$A_{ij} = \varepsilon^2 \left(\nabla \lambda^j, \nabla \lambda^i\right) + (M^\top M)_{ij}, \ \ B_{ij} = -\left(\lambda^j, \lambda^i\right), \ \ C_{ij} = \tau \left(L \nabla \lambda^j, \nabla \lambda^i\right),$$

where the definition

$$M_{i,j} = \varepsilon \left(\int_\Omega \lambda^j \right)_i, \qquad i = 1, \dots, N, \quad j = 1, \dots n, \tag{3.14}$$

of $M = (M_{ij}) \in \mathbb{R}^{N,n}$ provides $M^\top M U \cdot V = \varepsilon^2 \int_\Omega u \, dx \cdot \int_\Omega v \, dx$. The algebraic representation

$$\varphi(V) = \sum_{i=1}^{n} \varphi_i(V_i), \quad \text{with} \quad \varphi_i(\xi) = \Phi_\theta(\xi) \int_\Omega \lambda^i \, dx \tag{3.15}$$

of the nonlinearity ϕ_θ satisfies $\varphi(V) = \phi_\theta(v)$ and the right-hand sides $f = (f_i)$, $g = (g_i) \in \mathbb{R}^n$ are given by

$$f_i = -\left(K u^{\text{old}}, \lambda^i\right) + \left(M^\top M \int_\Omega u^{\text{old}} \, dx \right)_i, \qquad g_i = -\left(u^{\text{old}}, \lambda^i\right). \tag{3.16}$$

In analogy to (3.3) the variational problem (3.12) can be reformulated as the saddle point problem

$$\mathcal{L}(U,W) = \inf_{V\in\mathbb{R}^n} \sup_{Z\in\mathcal{S}^N} \mathcal{L}(V,Z) \tag{3.17}$$

for the Lagrangian

$$\mathcal{L}(U,W) = \frac{1}{2}AU\cdot U - f\cdot U + \varphi(U) + (BU-g)\cdot W - \frac{1}{2}CW\cdot W.$$

The construction and convergence analysis of nonsmooth Schur–Newton methods to be presented below will rely on the following reformulation.

Proposition 3.3. *The discrete spatial problem (VI) is equivalent to the set-valued saddle point problem*

(VIA) Find U, $W\in\mathbb{R}^n$ such that

$$\begin{pmatrix} A+\partial\varphi & B^T \\ B & -C \end{pmatrix}\begin{pmatrix} U \\ W \end{pmatrix} \ni \begin{pmatrix} f \\ g \end{pmatrix}$$

with the symmetric, positive definite matrix $A\in\mathbb{R}^{n,n}$, $B\in\mathbb{R}^{n,n}$, the symmetric, positive semi-definite matrix $C\in\mathbb{R}^{n,n}$, the subdifferential $\partial\varphi$ of the lower semi-continuous, proper convex functional $\varphi:\mathbb{R}^n\to\mathbb{R}$ and f, $g\in\mathbb{R}^n$ given in (3.13) – (3.16).

4. Nonsmooth Schur–Newton methods

In this section we consider the efficient algebraic solution of set-valued saddle point problems of the form (VIA) with a symmetric, positive definite matrix $A\in\mathbb{R}^{n,n}$, some matrix $B\in\mathbb{R}^{n,n}$, a symmetric, positive semi-definite matrix $C\in\mathbb{R}^{n,n}$, the subdifferential $\partial\varphi$ of a lower semi-continuous, proper convex functional $\varphi:\mathbb{R}^n\to\mathbb{R}$, and f, $g\in\mathbb{R}^n$ by nonsmooth Schur–Newton methods. This approach has been introduced and applied to discretized binary Cahn–Hilliard equations with obstacle potential in [22]. It was extended to more general nonsmooth nonlinearities and applied to a binary Cahn–Hilliard equation with logarithmic potential in [19, 20, 26]. For completeness, we present the basic ideas and convergence results, referring to [20, 22] for details.

4.1. Nonlinear Schur complement and unconstrained minimization.

Nonsmooth Schur–Newton methods are based on the reformulation of the set-valued saddle point problem (VIA) as a dual, unconstrained minimization problem. In a first step, we eliminate the primal variable U from (VIA).

Lemma 4.1. *The set-valued saddle point problem (VIA) is equivalent to the nonlinear system*

$$H(W) = 0 \tag{4.1}$$

with the single valued, Lipschitz continuous nonlinear Schur complement

$$H(W) = -B(A+\partial\varphi)^{-1}(f-B^\top W) + CW + g \tag{4.2}$$

in the sense that (U,W) is a solution of (VIA) if and only if W solves (4.1) *and $U=(A+\partial\varphi)^{-1}(f-B^TW)$.*

Proof. The inverse $(A+\partial\varphi)^{-1}$ of $A+\partial\varphi$ is single valued and Lipschitz continuous with Lipschitz constant given by the inverse of the coercivity constant of A, because A is s.p.d. and φ is lower semi-continuous, proper convex [14, Part One, Chapter II]. Now the assertion follows from straightforward computation. □

In the linear case $\partial\varphi \equiv 0$, it is well-known that the Schur complement $BA^{-1}B^\top + C$ is symmetric and positive definite. We now provide an extension of this property to the present nonlinear case.

Proposition 4.1. *There is a Fréchet-differentiable convex functional $h : \mathbb{R}^n \to \mathbb{R}$ such that $H = \nabla h$.*

Proof. Using Corollary 5.2 in [13, p. 22], it follows that $H = \partial h$ is the subdifferential of

$$h(W) = -\inf_{V\in\mathbb{R}^n} \mathcal{L}(V,W) = -\mathcal{L}((A+\partial\varphi)^{-1}(f-B^TW),W).$$

As $\partial h = H$ is single-valued and continuous, h is even Fréchet-differentiable and $H = \nabla h$ is the Fréchet derivative of h. □

As a direct consequence of Lemma 4.1 and Proposition 4.1, the set-valued saddle point problem (VIA) is equivalent to find $W \in \mathbb{R}^n$ such that

$$h(W) \le h(V) \qquad \forall V \in \mathbb{R}^n \tag{4.3}$$

and then solve $AU + \partial\varphi(U) \ni f - B^TW$. We emphasize that (4.3) now is an unconstrained convex minimization problem for an LC^1 function to which classical gradient-related descent methods can be applied.

4.2. **Gradient-related descent methods.** We give a short summary of this approach referring to text books like, e.g., [37] for the general theory or to [19, 22] for a more specific presentation.

We consider iterative methods for the approximation of minimizers of a given functional $h : \mathbb{R}^n \to \mathbb{R}$ of the form

$$W^{\nu+1} = W^\nu + \rho_\nu D^\nu. \tag{4.4}$$

The search directions $D^\nu \in \mathbb{R}^n$ are called *gradient-related*, if for any sequence $(W^\nu) \subset \mathbb{R}^n$ the conditions

$$\nabla h(W^\nu) = 0 \quad \Longleftrightarrow \quad D^\nu = 0$$

and

$$-\nabla h(W^\nu)\cdot D^\nu \ge c_D|\nabla h(W^\nu)||D^\nu|$$

hold for all $\nu \in \mathbb{N}$ with a constant $c_D > 0$ independent of ν. For example, the gradients $D^\nu = -\nabla h(W^\nu)$ are gradient-related, and we obtain the classical gradient method for this choice. Faster convergence speed can be expected for preconditioned gradient methods as resulting from search directions of the form

$$D^\nu = -S_\nu^{-1}\nabla h(W^\nu) \tag{4.5}$$

with a sequence $(S_\nu) \subset \mathbb{R}^{n,n}$ of suitable symmetric, positive definite preconditioners. Such search directions are gradient-related, if there are constants $\gamma, \Gamma > 0$ such that

$$\gamma|V|^2 \le S_\nu V\cdot V \le \Gamma|V|^2 \tag{4.6}$$

holds uniformly in $\nu \in \mathbb{N}$.

While the search directions D^ν are constructed to allow for suitable descent of the functional h, the step sizes ρ_ν should guarantee that this descent is actually realized. A sequence $(\rho_\nu) \subset \mathbb{R}$ of step sizes is called *efficient*, if

$$h(W^\nu + \rho_\nu D^\nu) \leq h(W^\nu) - c_S \left(\frac{\nabla h(W^\nu) \cdot D^\nu}{|D^\nu|} \right)^2 \tag{4.7}$$

holds with a constant $c_S > 0$ independent of ν.

Theorem 4.1. *Assume that the search directions take the form $D^\nu = -S_\nu^{-1} \nabla h(W^\nu)$ with symmetric, positive definite preconditioners $S_\nu \in \mathbb{R}^{n,n}$ satisfying (4.6), that the step sizes ρ_ν are efficient in the sense of (4.7), and that h has a unique minimizer. Then the sequence (W^ν) produced by (4.4) converges to the minimizer of h for $\nu \to \infty$.*

Proof. See Theorem 5.2 and Theorem 5.7 in [19]. The proof presented there is based on the fact that uniqueness of the minimizer implies compactness of the sublevel set $\{W \in \mathbb{R}^n \mid h(W) \leq h(W^0)\}$. Using this the rest can essentially be shown with standard arguments as, e.g. in, [37]. □

Efficiency of the step sizes ρ_ν can be guaranteed by various strategies such as, e.g., the Armijo rule (see, e.g., [11, Chapter 3] for a detailed discussion). In order to reduce the number of tuning parameters involved, we propose a strategy that approximates the minimizer of h along $W^\nu + \rho D^\nu$, $\rho \in \mathbb{R}$.

Proposition 4.2. *Assume that the search directions in the descent method (4.4) take the form $D^\nu = -S_\nu^{-1} \nabla h(W^\nu)$ and that*

$$\nabla h(W^\nu + \rho_\nu D^\nu) \cdot D^\nu \in [\alpha \nabla h(W^\nu) \cdot D^\nu, 0]$$

holds for all $\nu \in \mathbb{N}$ with some $\alpha \in [0, 1)$ independent of ν. Then the step sizes ρ_ν are efficient.

Proof. See Proposition 5.4 in [19]. □

Utilizing Proposition 4.2 with fixed $\alpha \in (0, 1)$, efficient step sizes ρ_ν can be computed by a simple bisection algorithm. However, as each iteration step requires the evaluation of $H = \nabla h$ and thus of $(A + \varphi)^{-1}$, this procedure might be quite costly. The actual computation of ρ_ν can be avoided, if the monotonicity test

$$|D^\nu| \leq \sigma |D^{\nu-1}| \tag{4.8}$$

holds with some fixed $\sigma < 1$. In this case, convergence is preserved for $\rho_\nu = 1$. We refer to [19, Theorem 5.4] for details.

Remark 4.1. *The above convergence results also remain valid, if the descent directions D^ν are replaced by approximations $\tilde{D}^\nu$ which are sufficiently accurate in the sense that the conditions*

$$\tilde{D}^\nu \cdot \nabla h(W^\nu) < 0, \qquad |D^\nu - \tilde{D}^\nu| / |\tilde{D}^\nu| \to 0$$

are satisfied. For a detailed analysis of such inexact versions we refer to [19].

In particular, Remark 4.1 allows for inexact evaluation of the preconditioner S_ν^{-1}.

4.3. **Nonsmooth Newton-like descent directions.** In the light of Theorem 4.1 the gradient-related descent method (4.4) with search directions of the form (4.5) converges globally for any sequence of symmetric positive definite preconditioners S_ν with the property (4.6). We now aim at selecting S_ν in such a way that the convergence is locally fast. For a sufficiently smooth functional h, the Jacobian $S_\nu = \nabla^2 h(W^\nu)$ would clearly be a desirable choice, because it leads to the classical Newton method with asmptotically quadratic convergence. Since for the given problem we cannot expect $\nabla h = H$ to be differentiable but only to be Lipschitz, the choice $S_\nu \in \partial_C H(W^\nu)$ with ∂_C denoting the generalized Jacobian in the sense of Clarke [10] would be a natural generalization. However, an element of $\partial_C H(W^\nu)$ is difficult to obtain, since, in general, we cannot make use of the chain rule. Following [19, 22], we will therefore construct related nonsmooth Newton-like preconditioners S_ν by *postulating* the chain rule. We will focus on the basic ideas of construction and present the resulting preconditioner for the given problem (4.1), referring to [19, 22] for details.

Our starting point is the observation is that some of the components of $(A + \partial\varphi)^{-1}$ are smooth in a given $Y \in \mathbb{R}^n$ while the others are not. To be precise, we introduce the inactive set

$$\bar{\mathcal{I}}(Y) := \{1 \le i \le n \mid \partial\varphi_i \text{ is single-valued and differentiable in } Y_i\}.$$

For the given φ defined in (3.15), we obtain

$$\bar{\mathcal{I}}(Y) = \{1 \le i \le n \mid Y_i > 0\}. \tag{4.9}$$

It turns out that the i-th component of the inverse $(A + \partial\varphi)^{-1}$ is differentiable in Y, if $i \in \bar{\mathcal{I}}((A + \partial\varphi)^{-1}Y)$. This observation motivates the linearization

$$\partial(A + \partial\varphi)^{-1}(Y) := (A + \varphi''(X))^+_{\bar{\mathcal{I}}(X)} \tag{4.10}$$

of $(A + \partial\varphi)^{-1}$ at a given $Y \in (A + \partial\varphi)(X)$. Here, $\varphi''(X)$ denotes the diagonal matrix with diagonal entries $\varphi_i''(X_i)$, the matrix $M^+ \in \mathbb{R}^{n,n}$ stands for the Moore–Penrose pseudoinverse of $M \in \mathbb{R}^{n,n}$, and $M_{\bar{\mathcal{I}}} \in \mathbb{R}^{n,n}$ denotes the so-called *truncated* matrix

$$(M_{\bar{\mathcal{I}}})_{i,j} = \begin{cases} M_{i,j} & \text{if } i,j \in \bar{\mathcal{I}}, \\ 0 & \text{else.} \end{cases}$$

Note that the multiplication of the matrix $(A + \varphi''(X))^+_{\bar{\mathcal{I}}(X)}$ with a vector amounts to the solution of a reduced linear system with a coefficient matrix consisting of the i-th row and column unit vector in $\mathbb{R}^n$, if $i \notin \bar{\mathcal{I}}(X)$ and remaining entries taken from $A + \varphi''(X)$, respectively.

The definition (4.9) of inactive sets is well-suited for the deep quench limit $\theta = 0$, because then the second derivatives of $\varphi_i(Y_i)$ are uniformly bounded, in fact equal to zero. This is different for the logarithmic potential, where the property $\varphi(Y_i) \to \infty$ for $Y_i \to 0$ might lead to badly scaled linearizations of the form (4.10). As a remedy we modify the definition of the active set according to

$$\mathcal{I}(Y) := \{1 \le i \le n \mid Y_i > \delta\} \tag{4.11}$$

with some fixed $\delta > 0$ such that $\varphi_i(Y_i) \le c_T$ holds with a corresponding fixed constant c_T. We will use $c_T = 10^8$ in our numerical computations. On this background,

we finally define the linearization

$$\partial(A+\partial\varphi)^{-1}(Y) = (A+\varphi''(X))^{+}_{\mathcal{I}(X)}, \qquad X = (A+\partial\varphi)^{-1}(Y),$$

of $(A+\partial\varphi)^{-1}$ at some given $Y \in \mathbb{R}^n$. Now, postulating the chain rule with $\partial(A+\partial\varphi)^{-1}$as inner derivative, we obtain the nonsmooth Newton-like linearization

$$\partial H(Y) = B(A+\varphi''(X))^{+}_{\mathcal{I}(X)}B^T + C, \qquad X = (A+\partial\varphi)^{-1}(f - B^T Y),$$

of H defined in (4.2) at some given $Y \in \mathbb{R}^n$.

The candidate $\partial H(W^\nu)$ for a preconditioner S_ν is symmetric and positive definite, if and only if $\mathcal{I}(U^\nu) \neq \emptyset$ holds with $U^\nu = (A+\partial\varphi)^{-1}(W^\nu)$, because we get $\partial H(W^\nu) = C$ otherwise, and C is only positive semi-definite. Hence, in the non-generic case $\mathcal{I}(U^\nu) = \emptyset$, we regularize $\partial H(W^\nu)$, e.g., by adding the scaled mass matrix τB to obtain

$$S_\nu = \begin{cases} \partial H(W^\nu) & \text{if } \mathcal{I}(U^\nu) \neq \emptyset, \\ \partial H(W^\nu) + \tau B & \text{else.} \end{cases} \tag{4.12}$$

Definition 4.1. *The gradient-related descent method* (4.4) *with search directions* $D^\nu = -S_\nu^{-1}H(W^\nu)$*, preconditioners* $S_\nu \in \mathbb{R}^{n,n}$ *defined in* (4.12)*, and efficient step sizes* ρ_ν *is called* nonsmooth Schur–Newton iteration *for the set-valued saddle point problem (VIA).*

Recall that efficient step sizes ρ_ν can be computed utilizing Proposition 4.2.

Theorem 4.2. *Assume that (VIA) has a unique solution* (U, W)*. Then, for any initial iterate* $W^0 \in \mathbb{R}^n$*, the nonsmooth Schur–Newton iteration converges to the solution* W *and* $U = (A+\partial\varphi)^{-1}(f - B^T W)$*.*

Proof. In the light of Theorem 4.1, it only remains to show that the preconditioners S_ν, defined in (4.12) are s.p.d. and satisfy condition (4.6). We refer to Theorem 5.7 in [19]. □

Recall that sufficient conditions for uniqueness are given in Theorem 3.2. Global convergence also holds for suitable inexact versions of (4.4) according to Remark 4.1.

By construction, we generally cannot expect $\partial H(W^\nu)$ to be contained in the set of generalized Jacobians in the sense of Clarke. Hence, the general theory of semi-smooth Newton methods cannot be applied to show local quadratic convergence. However, exploiting that the underlying solution space is finite dimensional, related results can be easily shown directly.

Remark 4.2. *Assume that the parameter* $\delta > 0$ *in* (4.11) *is sufficiently small and that the monotonicity test in* (4.8) *is not used. Then the nonsmooth Schur–Newton method as applied to the set-valued saddle point problem (VIA) locally reduces to a classical Newton iteration in case of the logarithmic potential* $\theta > 0$*, and is even locally exact in the deep quench limit* $\theta = 0$*.*

The numerical relevance of these asymptotic results is limited: For $\theta > 0$ sufficiently small parameters $\delta > 0$ typically lead to severe ill-conditioning of the arising Hessian matrices and to convergence radii of the Newton iteration that are smaller than machine precision.

The general convergence analysis of nonsmooth Schur–Newton methods is based on arguments restricted to finite dimensional spaces (see [19, 22] for a detailed discussion). Convergence results in function spaces are available only in special

cases [29]. However, numerical computations indicate mesh-independent convergence speed for initial iterates provided by nested iteration (cf. [19, 22] and the numerical experiments to be reported below). Theoretical validation of such kind of local mesh independence is still open.

4.4. **Algorithmic aspects.** Rewriting the nonsmooth Schur–Newton iteration introduced in Definition 4.1 in primal–dual form

$$U^\nu = (A + \partial\varphi)^{-1}(f - B^T W^\nu)$$
$$W^{\nu+1} = W^\nu + \rho_\nu S_\nu^{-1}(BU^\nu + CW^\nu + g),$$

we obtain a preconditioned Uzawa method. Each iteration step amounts to the update of the primal variable U^ν, the evaluation of the preconditioned residual $S_\nu^{-1}(BU^\nu + CW^\nu + g)$, and the selection of a suitable step size ρ_ν.

The first substep is equivalent to the solution of the minimization problem

$$U^\nu = \arg\min_{V\in\mathbb{R}^n} \tfrac{1}{2}AV\cdot V + \varphi(V) - (f - B^T W^\nu)\cdot V. \tag{4.13}$$

While there is a vast literature about elliptic obstacle problems emerging in the deep quench limit $\theta = 0$ (cf., e.g., [12, 21] and the references cited therein), fast solvers for the logarithmic potential $\theta > 0$ that show a robust convergence behavior for $\theta \to 0$ are still rare (see, however, [32, 33]). In the numerical experiments to be reported below, we apply the *truncated nonsmooth Newton method* (TNNMG) [19, 21, 24] that combines robustness for $\theta \to 0$ with similar efficiency as observed for classical multigrid methods in the linear self-adjoint case. Note that optimal complexity of each iteration step even for the dense matrix A is achieved by exploiting that A is the sum of a sparse matrix and a dense low-rank matrix with a known product representation [17].

The preconditioned residual $W^{\nu+1/2} = S_\nu^{-1}(BU^\nu + CW^\nu + g)$ can be computed by (approximately) solving a truncated linear saddle point problem of the form

$$\begin{pmatrix} \hat{A} & \hat{B}^T \\ \hat{B} & -\hat{C} \end{pmatrix} \begin{pmatrix} \hat{U} \\ W^{\nu+1/2} \end{pmatrix} = \begin{pmatrix} \hat{f} \\ g \end{pmatrix} \tag{4.14}$$

with the symmetric, positive definite matrix $\hat{A} \in \mathbb{R}^{\hat{n},\hat{n}}$, $\hat{n} = n - \#\mathcal{I}(U^\nu)$, obtained by eliminating the i-th row and column of A for all $i \in \mathcal{I}(U^\nu)$, the matrix $\hat{B} \in \mathbb{R}^{\hat{n},n}$ obtained by eliminating the i-th row of B for all $i \in \mathcal{I}(U^\nu)$, and $\hat{C} = C$, if $\hat{n} > 0$ or $C = C + \tau B$ otherwise. In the numerical experiments to be reported below, we use a preconditioned GMRES iteration with a truncated version of the multigrid method with successive Vanka smoother suggested by Schöberl and Zulehner [38] as a preconditioner. For an overview on other methods for the numerical solution of linear saddle point problems, we refer to [5].

Efficient step sizes ρ^ν can be computed utilizing Proposition 4.2 which requires the evaluation of $\nabla h(W^\nu + \rho D^\nu) = H(W^\nu + \rho D^\nu)$ and thus the solution of a minimization problem of the form (4.13) in each bisection step. Recall that this costly procedure can be avoided for iterates that are sufficiently accurate in the sense that the monotonicity test (4.8) is passed.

5. Numerical experiments

5.1. **Problem, discretization and subproblem solvers.** We consider the vector-valued Cahn-Hilliard equation (2.5) with $L = P = I - \frac{1}{N}(\mathbf{1},\dots,\mathbf{1})$, $\varepsilon^2 = 5\cdot 10^{-3}$ and

logarithmic potential Ψ_θ defined by (2.2), where K is given by (2.4) with $\theta_c = 1.0$ on the computational domain $\Omega = (-1,1)\times(-1,1)$. We select $N = 4$ components and the temperature $\theta = 0.1$, if not stated otherwise. To obtain initial conditions with similar granularity for varying N, 200 circles with radius $0.1 - 0.15$ are randomly distributed over Ω and randomly assigned to the different components.

Throughout the following, we use the uniform time step size $\tau = 10^{-3}$ and a grid hierarchy $\mathcal{T}_1, \dots, \mathcal{T}_j$ as obtained by successive refinement of the initial triangulation $\mathcal{T}_0$ consisting of two triangles with hypotenuse oriented from the upper left to lower right vertex of Ω. Though adaptivity is clearly mandatory in practical applications and would not affect any of the algorithms or results presented above, we assume for simplicity that the triangulations are uniformly refined, i.e., the edges of all triangles are bisected in each refinement step. If not stated otherwise, we select $j = 8$ refinement steps, providing the triangulation $\mathcal{T}_8$ with $n_8 = 66049$ vertices and the mesh size $h_j = 2^{-7} \approx 9\varepsilon$. In this case, we found that the interface is resolved by about 11 grid points.

For the iterative solution of the resulting algebraic subproblems, we consider the nonsmooth Schur-Newton method (NSNMG) presented in Section 4 with multigrid solution of the nonlinear non-smooth subproblems (4.13) and a preconditioned GMRES iteration for the linearized saddle point problems (4.14).

More precisely, non-smooth subproblems (4.13) are solved by a truncated non-smooth Newton multigrid method (TNNMG) [19, 21, 24]. Throughout the following, the iteration is executed almost up to machine precision, i.e., we use the stopping criterion

$$\|U^{\nu,k+1} - U^{\nu,k}\|_A < 10^{-13} \tag{5.1}$$

for the iterates $U^{\nu,k}$, $k = 1, \dots$, with $\|\cdot\|_A$ denoting the energy norm induced by the matrix A.

The linear saddle point problems (4.14) are solved by a preconditioned GMRES iteration with restart after 50 steps. The preconditioner is based on a straightforward extension of a truncated multigrid method with block-Gauß-Seidel smoother as suggested in [32, 38] to the vector-valued case. In the light of Remark 4.1 the iteration is stopped, if the ratio of the Euclidean norms of the preconditioned residual and the actual iterate is less than $\min\left(\zeta_1^\nu,\ \zeta_2\|W^\nu - W^{\nu-1}\|_{C,B}^2\right)$. Here, we chose $\zeta_1 = 10^{-1}$ and $\zeta_2 = 10^{-2}$ and the corrections $W^\nu - W^{\nu-1}$ of the overall NSNMG iteration are measured in the norm

$$\|V\|_{C,B} = \|V\|_C + \tau\|V\|_B, \qquad V \in \mathbb{R}^n, \tag{5.2}$$

generated by the positive semi-definite matrix C and the mass matrix B defined in (3.13).

The step sizes ρ^ν are computed according to Proposition 4.2.

The overall NSNMG iteration is terminated once its target, the dual variable W, is approximated sufficiently well, i.e. once the stopping criterion

$$\|W^{\nu+1} - W^\nu\|_{C,B} < \kappa 10^{-11}. \tag{5.3}$$

is satisfied with some $\kappa > 0$. We chose the default value $\kappa = 1$, if not stated otherwise.

5.2. **Evolution and distribution of computational work.** In our first experiment, we consider the evolution of $N = 6$ components Here, we chose $\kappa = 2$ in

the stopping criterion (5.3) in order to avoid the influence of round-off errors in our linear saddle point solver.

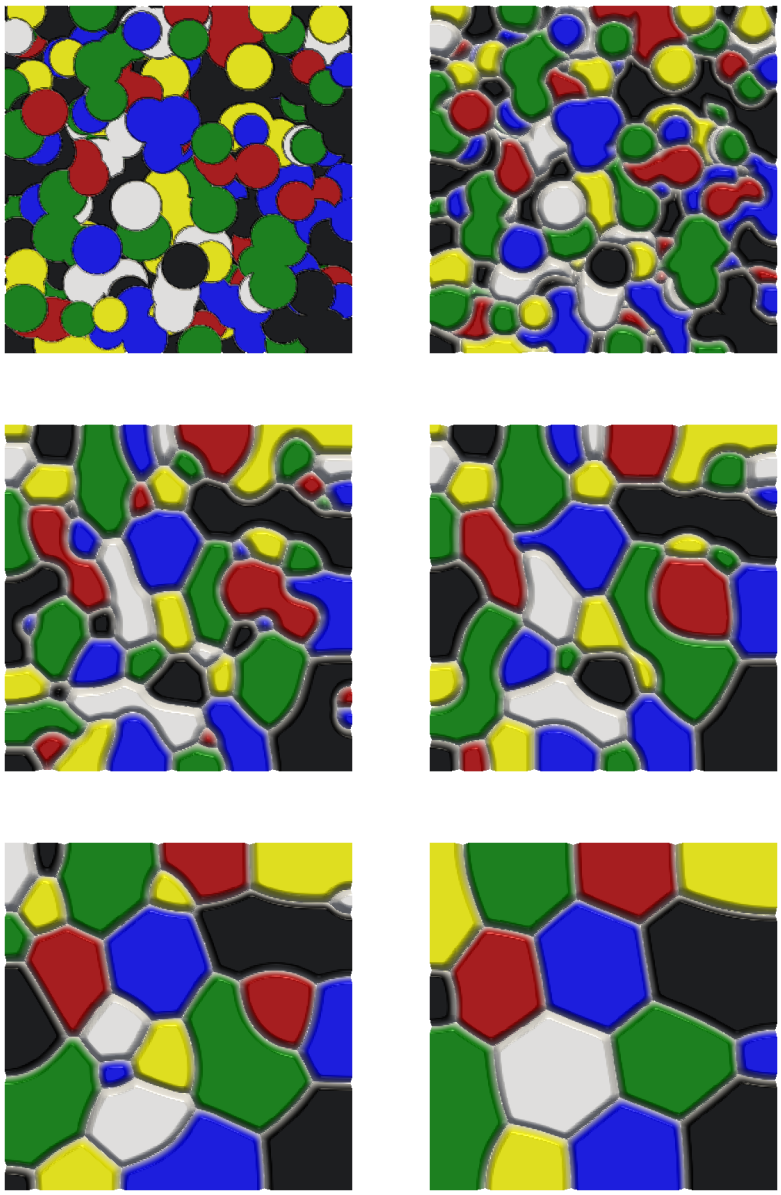

FIGURE 5.1. Initial condition u_0 and approximate order parameter $u(\cdot,t)$ at time $t = 1\tau,\ 20\tau,\ 50\tau,\ 200\tau,\ 1000\tau$.

The evolution over 1000 time steps is illustrated in Figure 5.1. As expected, we observe fast separation in the beginning and slower dynamics in course of the evolution. Triple, quadruple, and even quintuple junctions occur with nicely equilibrated

angles. It is also interesting to see that the evolution tends to a hexagonal structure of grains with equilibrated mass. Mass conservation is fulfilled up to 0.0053% over all time steps, which is in good accordance with our prescribed algebraic accuracy.

To illustrate the the amount of computational work, Figure 5.2 shows the total number of iterations by NSNMG (red), TNNMG (blue), and preconditioned GMRES (green) for each spatial problem, scaled by their respective values in the first timestep (6 (NSNMG), 45 (TNNMG), and 281 (preconditioned GMRES)), over the number of time steps. The initial iterate is obtained by nested iteration in each case. We observe exactly 6 NSNMG iterations for all spatial problems and

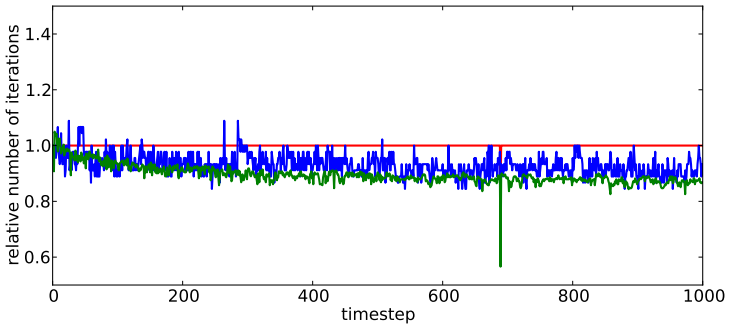

FIGURE 5.2. Total number of iterations by NSNMG (red), TNNMG (blue), and preconditioned GMRES (green), scaled by their value in the first timestep 6 (NSNMG), 45 (TNNMG), and 281 (preconditioned GMRES), over the number of time steps.

only slight changes in the performance of TNNMG and preconditioned GMRES. No damping was needed throughout the evolution. Hence, the solution of each subproblem (4.13) only required about 7 iterations of TNNMG. This is in accordance with previous computations, where TNNMG exhibited linear multigrid efficiency and mesh-independent convergence rates for initial iterates provided by nested iteration [19, 24]. The preconditioned GMRES needed more than 40 iterations for each linear solve and thus strongly dominates the overall computational work. Moreover, we found that our straightforward multigrid preconditioning did not provide mesh independence. Hence, the overall efficiency of NSNMG will benefit from more sophisticated linear saddle point solvers as have been studied elsewhere (see, e.g.,[36]).

5.3. **Influence of initial iterate, temperature, number of components, and spatial mesh size on the convergence speed.** In our next experiment, we come back to $N = 4$ components and study the influence of initial iterates W^0 and temperature θ on the convergence speed of NSNMG. As the performance of NSNMG hardly changed for different spatial problems (cf. Section 5.2), we concentrate on the first one. Figure 5.3 shows the approximate algebraic error $\|W^\nu - W^{\nu-1}\|_{C,B}$ over the number of NSNMG iterations for the temperatures $\theta = 0.5$ (blue), 0.1 (black), 0.001 (green), 0 (red) with "bad" initial iterates $W^0 = 0$ (dashed lines) and "good" initial iterates obtained by nested iteration (solid lines). For bad initial iterates, the iteration history can be separated into an asymptotic phase with slow convergence and step sizes $\rho^\nu < 1$ and into an asymptotic phase with superlinear convergence speed. The asymptotic phase is entered immediately for initial

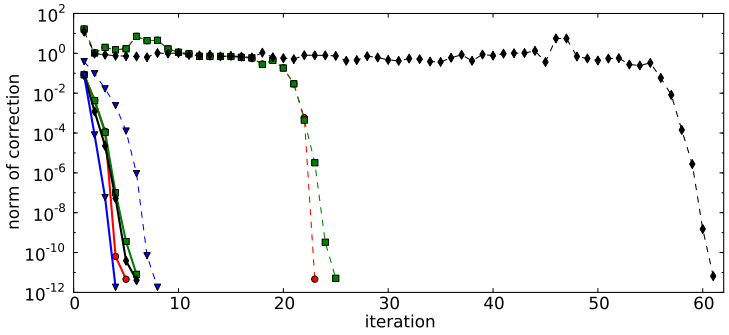

FIGURE 5.3. Approximate algebraic error $\|W^\nu - W^{\nu-1}\|_{C,B}$ over the number of NSNMG iterations for the temperatures $\theta = 0.0$ (red •), 0.001 (green ■), 0.1 (black ♦), 0.5 (blue ▼) with initial iterates $W^0 = 0$ (dashed lines) and nested iteration (solid lines) for the first spatial problem.

iterates obtained by nested iteration. While we observe a strong influence of the temperature θ on the duration of the asymptotic phase, it hardly seems to affect the asymptotic superlinear convergence speed.

This suggests robustness of NSNMG with respect to temperature θ for initial iterates as obtained by nested iteration, which is confirmed by our next experiment, as illustrated in Figure 5.4. The left picture shows the number of iterations as required to meet the stopping criterion (5.3) with $\kappa = 1$ over the inverse temperature $1/\theta$. We chose the values $\theta = i \cdot 10^{-1}$, $i = 1, \dots 9$, and $\theta = 10^{-i}$, $i = 2, \dots 10$. The

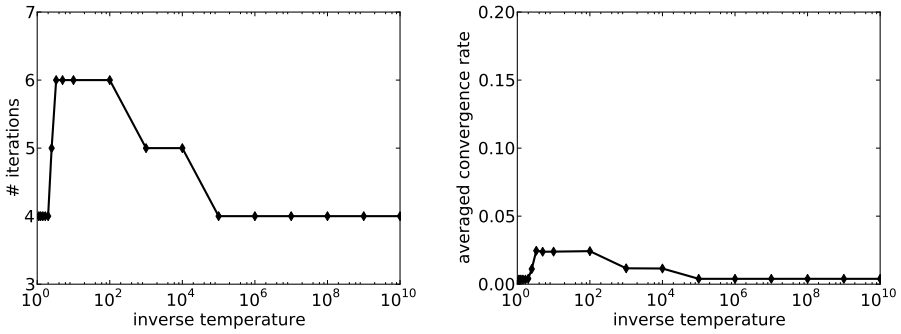

FIGURE 5.4. Number of NSNMG iterations over inverse temperature $1/\theta$ (left) and averaged convergence rates of NSNMG over $1/\theta$ (right).

corresponding averaged convergence rates are shown in the picture on the right. In a sense, problems with $\theta \approx \theta_c = 1$ and $\theta \approx 0$ seem to be a little bit easier to solve than problems with medium temperatures, such as, e.g. $\theta = 0.1$. Observe that the convergence behavior for $\theta = 10^{-5}$ can not be distinguished from the deep quench limit $\theta = 0$. At most 6 NSNMG iterations were required to reduce the approximate error by 10 orders of magnitude. The averaged NSNMG convergence rate is always

far below 0.05 so that usually one or two NSNMG steps would be enough to reduce the algebraic error below discretization accuracy. Each NSNMG step amounts to the total number of at most 23 TNNMG iterations and one inexact linear saddle point solution. This means that the average number of TNNMG iterations for each occurring subproblem (4.13) is less than 6, which nicely confirms efficiency and robustness of this method [19, 24]. Again the overall computational work is strongly dominated by the inexact linear saddle point solution which is partly due to the larger number of unknowns.

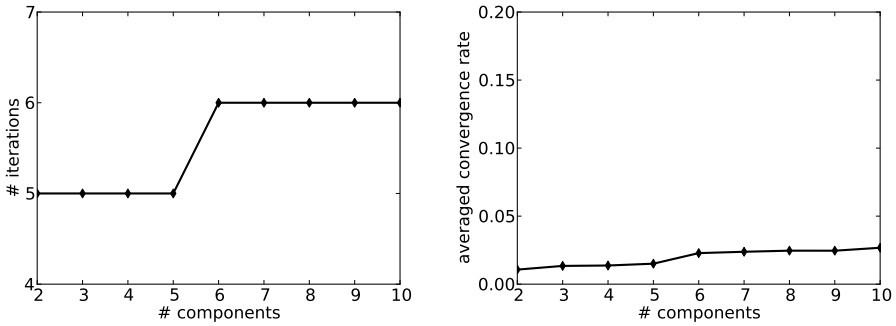

FIGURE 5.5. Number of NSNMG iterations over the number N of components (left) and averaged convergence rates of NSNMG over N (right).

In the next experiment, we assess the influence of the number N of components on the convergence speed of NSNMG. Here, we had to chose $\kappa = 7$ in the stopping criterion (5.3) in order to avoid the influence of round-off errors in our linear saddle point solver. The left picture of Figure 5.5 shows the number of NSNMG iterations over N, while the right picture shows the corresponding averaged convergence rates. Again, the initial iterates are obtained by nested iteration. The number of NSNMG

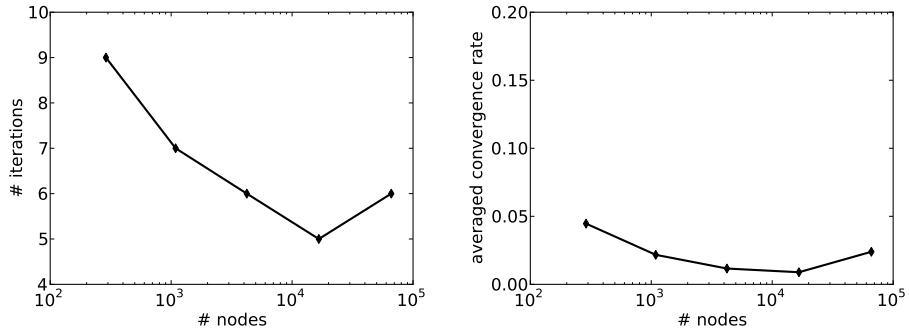

FIGURE 5.6. Number of NSNMG iterations over the number of vertices n_j on the refinement levels $j = 4 \dots, 8$, (left) and averaged convergence rates of NSNMG over n_j , $j = 4 \dots, 8$, (right).

iterations is varying between 5 and 6 over $N = 2, \dots, 10$ components, indicating considerable robustness of the convergence speed of NSNMG with respect to the

number of components. This robustness is preserved by TNNMG, which required less than a total number of 45 TNNMG iterations in each NSNMG step to solve the nonlinear nonsmooth subproblems (4.13) almost up to machine accuracy.

As the convergence theory presented in Section 4.2 is partly based on arguments that are restricted to finite dimensional spaces, we now investigate the mesh dependence of NSNMG. As we are interested in the local asymptotic convergence speed, the initial iterates are obtained by nested iteration. The left picture of Figure 5.6 shows the number of NSNMG iterations over the number n_j of vertices of the triangulations $\mathcal{T}_j$ on the levels $j = 4, \dots, 8$, while the right picture shows the corresponding averaged convergence rates. For the mesh size ranging from $h_4 = 2^{-3}$ to $h_8 = 2^{-7}$, the number of NSNMG iterations is bounded by 9 suggesting local mesh-independent convergence of NSNMG and even local convergence of a related approach in function space (see [29] for a first result in this direction). Theoretical justification will be the subject of future research.

References

[1] J. W. Barrett and J. F. Blowey. An error bound for the finite element approximation of a model for phase separation of a multi-component alloy. *IMA J. Numer. Anal.*, 16:257–287, 1996.

[2] J. W. Barrett and J. F. Blowey. Finite element approximation of a model for phase separation of a multi-component alloy with non-smooth free energy. *Numer. Math.*, 77:1–34, 1997.

[3] J. W. Barrett and J. F. Blowey. An improved error bound for a finite element approximation of a model for phase separation of a multi-component alloy with a concentration dependent mobility matrix. *Numer. Math.*, 88:255–297, 2001.

[4] J.W. Barrett and J.F. Blowey. An optimal error bound for a finite element approximation of a model for phase separation of a multi-component alloy with non-smooth free energy. *M2AN, Math. Model. Numer. Anal.*, 33(5):971–987, 1999.

[5] M. Benzi, G. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.

[6] J.F. Blowey, M.I.M. Copetti, and C.M. Elliott. Numerical analysis of a model for phase separation of a multi-component alloy. *IMA J. Numer. Anal.*, 16(1):111–139, 1996.

[7] J.F. Blowey and C.M. Elliott. The Cahn-Hilliard gradient theory for phase separation with non-smooth free energy part ii: Numerical analysis. *Euro. Jn. Appl. Math.*, 3:147–179, 1993.

[8] P. Boynova and M. Neytcheva. Efficient numerical solution of discrete multi-component Cahn-Hilliard systems. Preprint 09, Department of Information Technology, Uppsala Universitet, 2012.

[9] L. Bronsard, H. Garcke, and B. Stoth. A multi-phase Mullins-Sekerka system: Matched asymptotic expansions and an implicit time discretisation for the geometric evolution problem. *Proc. R. Soc. Edinb., Sect. A, Math.*, 128(3):481–506, 1998.

[10] F. H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley, New York, 1983.

[11] P. Deuflhard. *Newton Methods for Nonlinear Problems*. Springer, Berlin, Heidelberg, 2004.

[12] Z. Dostál. *Optimal Quadratic Programming Algorithms*. Springer, 2009.

[13] I. Ekeland and R. Temam. *Convex Analysis*. Studies in Mathematics and its Applications. North-Holland, Amsterdam New York Oxford, 1976.
[14] I. Ekeland and R. Temam. *Convex Analysis*. North-Holland, 1976.
[15] C.M. Elliott and S. Luckhaus. A generalised diffusion equation for phase separation of a multicomponent mixture with interfacial free energy. Preprint 887, IMA, University of Minnesota, 1991.
[16] D. De Fontaine. An analysis of clustering and ordering in multicomponent solid solutions I. Stability criteria. *J. Phys. Chem. Solids*, 33:287–310, 1972.
[17] C. Gräser. Analysis und Approximation der Cahn-Hilliard Gleichung mit Hindernispotential. Diplomarbeit, Freie Universität Berlin, 2004.
[18] C. Gräser. Globalization of nonsmooth Newton methods for optimal control problems. In K. Kunisch, G. Of, and O. Steinbach, editors, *Numerical Mathematics and Advanced Applications, Proceedings of ENUMATH 2007*, pages 605–612. Springer, 2008.
[19] C. Gräser. *Convex Minimization and Phase Field Models*. PhD thesis, Freie Universität Berlin, 2011.
[20] C. Gräser. Nonsmooth Schur–Newton methods for nonsmooth saddle point problems. in preparation, Matheon Berlin, 2013.
[21] C. Gräser and R. Kornhuber. Multigrid methods for obstacle problems. *J. Comp. Math.*, 27(1):1–44, 2009.
[22] C. Gräser and R. Kornhuber. Nonsmooth Newton methods for set-valued saddle point problems. *SIAM J. Numer. Anal.*, 47(2):1251–1273, 2009.
[23] C. Gräser, R. Kornhuber, and U. Sack. Time discretizations of anisotropic Allen–Cahn equations. *IMA J. Numer. Anal.*, to appear.
[24] C. Gräser, U. Sack, and O. Sander. Truncated nonsmooth Newton multigrid methods for convex minimization problems. In M. Bercovier, M. Gander, R. Kornhuber, and O. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, LNCSE, pages 129–136. Springer, 2009.
[25] C. Gräser and O. Sander. The dune-subgrid module and some applications. *Computing*, 8(4):269–290, 2009.
[26] C. Gräser and O. Sander. Polyhedral Gauss–Seidel converges. *SIOPT*, submitted.
[27] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer, Berlin, 1985.
[28] M Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13:865888, 2003.
[29] M. Hinze and M. Vierling. Variational discretization and semi-smooth Newton methods; implementation, convergence and globalization in pde constrained optimzation with control constraints. *Optim. Meth. Software*, page to appear, 2012.
[30] V. Styles J. Barrett, R. Nürnberg. Finite element approximation of a phase field model for void electromigration. *SIAM J. Numer. Anal.*, 42:738–772, 2004.
[31] J. Kim and K. Kang. A numerical method for a ternary Cahn–Hilliard system with degenerate mobility. *Appl. Numer. Math.*, 59:1029–1042, 2009.
[32] R. Kornhuber. On constrained Newton linearization and multigrid for variational inequalities. *Numer. Math.*, 91:699–721, 2002.

[33] R. Kornhuber and R. Krause. Robust multigrid methods for vector-valued Allen–Cahn equations with logarithmic free energy. *Comp. Visual. Sci.*, 9:103–116, 2006.
[34] P.L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *Siam J. Numer. Anal.*, 16:964–979, 1979.
[35] J.E. Morral and J.E. Cahn. Spinodal decomposition in ternary systems. *Acta Met.*, 19:1037, 1971.
[36] M.F. Murphy, G.H. Golub, and A. Wathen. A note on preconditioning for indefinite linear systems. *SIAM J. Sci. Comput.*, 21:1969–1972, 2000.
[37] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.
[38] J. Schöberl and W. Zulehner. On additive Schwarz-type smoothers for saddle point problems. *Numer. Math.*, 95(2):377–399, 2003.
[39] I. Steinbach, F. Pezolla, B. Nestler, M. Seeßelberg, R. Prieler, G.J. Schmitz, and J.L.L. Rezende. A phase-field concept for multiphase systems. *Physica D*, 94:135–147, 1996.

Carsten Gräser, Freie Universität Berlin, Institut für Mathematik, Arnimallee 6, D - 14195 Berlin, Germany
E-mail address: graeser@math.fu-berlin.de

Ralf Kornhuber, Freie Universität Berlin, Institut für Mathematik, Arnimallee 6, D - 14195 Berlin, Germany
E-mail address: kornhuber@math.fu-berlin.de

Uli Sack, Freie Universität Berlin, Institut für Mathematik, Arnimallee 6, D - 14195 Berlin, Germany
E-mail address: usack@math.fu-berlin.de